

Feature predictability flexibly supports auditory stream segregation or integration

Alexandra Bendixen^{1,2}, Susan L. Denham³, and István Winkler^{4,5}

¹Department of Psychology, Cluster of Excellence “Hearing4all”, European Medical School, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany

²Institute of Psychology, University of Leipzig, Leipzig, Germany

³Cognition Institute and School of Psychology, University of Plymouth, Plymouth, UK

⁴Institute of Psychology and Cognitive Neuroscience, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary

⁵Institute of Psychology, University of Szeged, Szeged, Hungary

Address correspondence to:

Alexandra Bendixen

Department of Psychology

Cluster of Excellence “Hearing4all”

European Medical School

Carl von Ossietzky University of Oldenburg

Ammerländer Heerstraße 114-118

D-26129 Oldenburg, Germany

Tel: +49-(0)441-798-4844

Fax: +49-(0)441-798-5522

E-mail: alexandra.bendixen@uni-oldenburg.de

Abstract

Many sound sources emit series of discrete sounds. Auditory perception must bind these sounds together (stream integration) while separating them from sounds emitted by other sources (stream segregation). One cue for identifying successive sounds that belong together is the predictability between their feature values. Previous studies have demonstrated that independent predictable patterns appearing separately in two interleaved sound sequences support perceptual segregation. The converse case, whether a joint predictable pattern in a mixture of interleaved sequences supports perceptual integration, has not yet been put to a rigorous empirical test. This was mainly due to difficulties in manipulating the predictability of the full sequence independently of the predictability of the interleaved subsequences. The present study implemented such an independent manipulation. Listeners continuously indicated whether they perceived a tone sequence as integrated or segregated, while predictable patterns set up to support one or the other percept were manipulated without the participants' knowledge. Perceptual reports demonstrate that predictability supports stream segregation or integration depending on the type of predictable pattern that is present in the sequence. The effects of predictability were so pronounced as to qualitatively flip perception from predominantly (62%) integrated to predominantly (73%) segregated. These results suggest that auditory perception flexibly responds to encountered regular patterns, favoring predictable perceptual organizations over unpredictable ones. Besides underlining the role of predictability as a cue within auditory scene analysis, the present design also provides a general framework that accommodates previous investigations focusing on sub-comparisons within the present set of experimental manipulations. Results of intermediate conditions shed light on why some previous studies have obtained little to no effects of predictability on auditory scene analysis.

Keywords

auditory scene analysis, stream segregation, stream integration, streaming, sound predictability, regular pattern, dynamic attending, perceptual bi-stability, perceptual switching

PACS classifications

43.66.Mk (Temporal and sequential aspects of hearing; auditory grouping in relation to music)

43.66.Ba (Models and theories of auditory processes)

43.66 Lj (Perceptual effects of sound)

43.66.Ki (Subjective tones)

43.66.Qp (Localization of sound sources)

Acknowledgment

This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG SCH 375/20-1, and DFG Cluster of Excellence 1077 “Hearing4all”), by the Hungarian Academy of Sciences (Magyar Tudományos Akadémia [MTA], Lendület project LP2012-36/2012 to I.W.), by the German Academic Exchange Service (Deutscher Akademischer Austauschdienst [DAAD], Project 56265741), and by the Hungarian Scholarship Board (Magyar Ösztöndíj Bizottság [MÖB], Project 39589). The experiment was realized using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN. The authors are grateful to Susann Duwe for collecting the data.

Introduction

The auditory system constantly needs to parse the acoustic environment into meaningful streams of information, such as the utterance of a speaker, the song of a bird, or the sound of a train moving on the rails. This parsing process is called *stream segregation* [1]. In turn, the individual elements that have been identified as belonging to one sound source (e.g., the speaker) need to be bound together over time (*stream integration*). These auditory scene analysis processes are supported by a number of acoustic grouping cues that can be summarized as *feature similarity* between the signals emitted by the different sound sources [2,3]. Signals that are similar to each other are likely to be integrated (i.e., perceived as originating from the same sound source), whereas dissimilar signals are likely to be segregated (i.e., perceived as originating from different sound sources). More recently, it has been shown that *feature predictability* is also taken into account within auditory scene analysis [4,5] (see also early work by [6]). Specifically, when the sounds within a stream are arranged in predictable patterns, listeners are more likely to segregate this stream from other acoustic input. The present study investigates whether predictable patterns can also be used as a grouping cue for stream integration.

Sequential auditory scene analysis is often investigated using the *auditory streaming paradigm* [7]. In this paradigm, participants are presented with a tone sequence made up of two different tones, ‘A’ and ‘B’. The tones are arranged in a continuously repeating ‘ABA–’ cycle, where ‘–’ represents a silent gap whose length is equal to the duration of the ‘A’ and ‘B’ tones. The difference between ‘A’ and ‘B’ lies in an acoustic feature such as pitch, location, timbre, or any combination thereof. These sequences can be perceived in at least two different ways. Depending on the amount of feature difference and the rate of tone presentation, participants tend to perceive the ‘ABA–’ sequence as originating from one sound source (*integrated*) with a galloping rhythm, or from two different sound sources (*segregated*), one made up of the ‘A’ tones, and the other made up of the ‘B’ tones repeating at half the rate of the ‘A’ tones [3,7]. With prolonged exposure to such sequences, perception switches back and forth between the integrated and segregated (and possibly

other) alternatives [8–15]. This perceptual bi- or multistability has been interpreted as a sign of the system exploring different grouping alternatives [16–18].

Early on, it was suggested that the perceptual decision between integration and segregation should not only depend on the feature (dis)similarity between the ‘A’ and ‘B’ tones, but also on regular (predictable) patterns formed by the tones within each putative stream [6,19]. This idea was based on the rationale that the signal emission pattern could be a qualifying characteristic of a given sound source (e.g., the pattern of a train moving on the rails). Effects of sequential predictability were indeed shown when participants were actively trying to hear out the predictable stream [20–22]. Because these effects pertained to temporal predictability, they were ascribed to rhythmic attending [6,19]. Attempts to demonstrate a more general role of sequential predictability within auditory scene analysis (without the involvement of attentional processes) initially proved unsuccessful [23,24]. Both studies showed that predictability does not affect the integrated or segregated perception of a sequence when listeners simply report their current percept, without any attempt to hear out a particular stream. The authors concluded that predictability is not used as an early, automatic grouping cue, but comes into play at a later stage of auditory scene analysis contingent upon the involvement of attentional processes (see also [1]).

More recently, the possible role of predictability within auditory scene analysis has been re-examined, partly motivated by a better understanding of the processing of auditory predictability [25–27]. The emerging consensus suggests that feature predictability is used as a cue for auditory stream segregation not only when a predictable stream is in the focus of attention [20–22,28] but also when a predictable stream needs to be suppressed in order to perform a given listening task (e.g., ignoring the repetitive sound of a train while following a conversation; see [4,28,29]). Feature predictability also biases perceptual reports with neutral listening instructions; i.e., when the listener does not attempt to hear a given sequence in any particular way [5,30,31]. Even during passive listening (i.e., without any active task associated with the sound sequence), predictability affects auditory scene analysis as shown by electroencephalographic measures [32].

This suggests that feature predictability is a more powerful grouping cue than previously assumed [1]. The precise mechanism by which predictability acts upon auditory scene analysis is still unclear. One interpretation that is consistent with the available evidence suggests that detecting the presence of a predictable pattern stabilizes the corresponding auditory stream by increasing the perceptual coherence of its elements, thus making this stream easier to maintain for longer periods of time [5,17,18,30]. If this interpretation is correct, then the same mechanism should apply when the listener perceives only one stream – in other words, predictable patterns should stabilize an integrated percept when compared to an arrangement of similar tones with randomly jittered feature values. This inference has not yet been tested empirically.

In the aforementioned studies [4,5,28–30,32], feature predictability was set up so that predictable patterns were introduced separately into the ‘A’ and/or ‘B’ tone sets (e.g., continuous repetition of ‘123123...’, with ‘1’, ‘2’ and ‘3’ representing three slightly different frequency values within the ‘A’ tone set). The resulting predictable condition was then compared with a condition in which the tones within each set were arranged with the same amount of feature variation but in random succession (e.g., ‘213312122313...’). As predictability was embedded separately into the ‘A’ and/or ‘B’ tone sets, the overall ‘ABA–’ sequence was never predictable. Predictability was hypothesized to support the segregated perceptual organization in this case, and indeed, an increase in measures of stream segregation was found [4,5,28–30,32].

One previous study attempted to address the converse case, a possible effect of feature predictability on stream integration [23]. This was done by introducing a predictable pattern into the whole (‘ABAB’) sequence, and contrasting this with a random arrangement of the sequence. Unfortunately, in this specific design, introducing predictability into the whole sequence also made the ‘A’ and ‘B’ subsequences predictable. The co-variation of the presence of predictability in the overall sequence (corresponding to the integrated percept) and in the subsequences (corresponding to the segregated percept) might have maintained the balance between the integrated and segregated perceptual organizations, causing the null effect observed by the authors. Indeed, the authors

mention this as a possible confound, and suggest that experiments with independent manipulations of the predictability of the overall sequence and the subsequences are needed [23]. To the best of our knowledge, no such independent manipulation has been performed up to now.

To study the effects of predictability on integration and segregation independently of each other, the present experiment provides an independent manipulation of predictability in the overall ‘ABA–’ sequence and the ‘A’ and ‘B’ subsequences. This was achieved by systematically changing the sequential dependencies between successive tones within and/or across the ‘ABA–’ triplets, such that these tones would or would not be predictive of each other. In the framework suggested by [18], this corresponds to strengthening or weakening the links between pairs of sounds, which would have immediate consequences on their perceptual coherence (see also [33]).

As in previous experiments studying the role of predictability on stream segregation (e.g. [5]), a moderate amount of feature jitter was introduced into the ‘A’ and ‘B’ tone sets to yield a certain degree of unpredictability as a baseline for assessing the effects of introducing predictability. The jitter pertained to the features of frequency, location, and onset time, which were conjointly varied to enhance the chances of predictability extraction.

--- insert Figure 1 at about here ---

Experimental conditions differed in how the feature values of successive tones related to each other (cf. Figure 1). In condition 1 (*no perceptual organization predictable*), the feature value for every single ‘A’ and ‘B’ tone was chosen randomly. This condition had a lower degree of predictability than any other condition and can thus be regarded as a baseline.

In condition 2 (*segregated organization predictable*), the feature values of successive tones within each subsequence (i.e., separately for the ‘A’ and ‘B’ tones) changed in a predictable manner. Based on the results of previous studies [5,30], this manipulation was hypothesized to bias perception towards stream segregation compared to the baseline condition 1.

In condition 3 (*integrated organization predictable*), all tones within an ‘ABA’ triplet shared the same feature value, whilst the feature values changed randomly from triplet to triplet. Thus all

the links connecting ‘A’ and ‘B’ tones would be strengthened, whereas none of the links between successive ‘B’ tones would be strengthened, and only half of the links connecting successive ‘A’ tones (the ones falling within but not the ones across triplets). Therefore, this manipulation was hypothesized to bias perception towards stream integration.

Condition 4 (*integrated and segregated organizations predictable*) combined the predictability manipulations implemented in conditions 2 and 3. Consequently, all possible sequential links were strengthened, causing the tone sequence to be predictable both within the integrated and within the segregated perceptual organizations. Perceptual reports in this condition were hypothesized to fall between those in conditions 2 and 3.

Finally, in condition 5 (*no-variation control*) all feature values were constant. This was done for comparison with the classical variant of the ‘ABA–’ streaming paradigm, which usually does not include any feature jitter. Note that in terms of predictability, condition 5 is most similar to 4 because in both conditions, feature values are predictable within the integrated as well as the segregated perceptual organization. A comparison of these conditions is informative about the impact of feature similarity within each stream independently of the impact of predictability. Note that feature predictability and similarity were confounded in some previous studies in which the predictable condition was created via feature constancy (as in the present condition 5) whereas the unpredictable condition was created via feature variation (as in condition 1). The present design thus allows us to disentangle the effects of feature similarity from those of predictability.

Methods

Participants. Twenty healthy volunteers (mean age 23.7 years, range 19-30 years; 2 left-handed, 1 ambidextrous, 17 right-handed; 8 male, 12 female) took part in the study. All participants had reportedly normal hearing. None of the participants were taking any medication affecting the central nervous system. In compliance with the Declaration of Helsinki, participants gave written

informed consent after the experimental procedures had been explained to them. Participants received course credit or modest financial compensation for their participation.

Apparatus and Stimuli. Participants were seated in an acoustically shielded chamber. Sinusoidal tones with a mean level of 70 dB sound pressure level were presented binaurally via Sennheiser HD25-1 closed-back on-ear headphones in a cyclically repeating ‘ABA–’ pattern. Stimulus levels were measured using a HEAD acoustics HMS III.0 artificial head measurement system. Participants were provided with a response keypad containing four buttons to be pressed with the middle and index fingers of their left and right hands.

Stimuli were arranged in five conditions defined by the presence or absence of jitter in the feature values (see below) and by four different versions of arranging the jitter (see below for detailed description, see Figure 1 for a schematic illustration for each condition). Each condition was administered in a separate 4-minute stimulus block.

Participants received further stimulus blocks with the same instructions and similar (‘ABA–’) stimuli. For half of the participants, the additional blocks with similar manipulations but different parameters were conducted before the current experimental blocks. For the other half of the participants, the additional blocks with a different manipulation were administered after the current experimental blocks. Results of these two pilot experiments are not reported here.

The ‘A’ and ‘B’ tones were delivered with a common duration of 75 ms (including raised cosine onset and offset ramps of 10 ms duration each). The mean frequency of the ‘A’ tones was 400 Hz, and the mean frequency of the ‘B’ tones was 5 semitones higher, i.e., 534 Hz. Depending on condition, both tone sets varied in frequency around their mean (nominal) value by an amount of ± 0.5 semitones. This small amount of variation was chosen to preserve similarity within each set while promoting a clear differentiation between the two sets.

Depending on condition, the perceived source location of the tones was also varied. This variation was implemented via jittering the interaural time differences (ITD values of ± 500 μ s) and the interaural level differences (ILD values of ± 10 dB) around the mean values of ITD=0 μ s,

ILD=0 dB (perceived as coming from the midline). ITD and ILD values were co-varying, intended to simulate tones coming approximately from $\pm 90^\circ$ left/right from the midline.

The ‘ABA–’ cycle was delivered at a stimulus onset asynchrony (SOA) of 150 ms between successive elements, thus giving the whole cycle a duration of 600 ms. Consecutive ‘A’ tones were separated by a 300 ms SOA with a ‘B’ tone inserted midway between every second pair of consecutive ‘A’ tones. Thus the SOA between successive ‘B’ tones was 600 ms. Depending on condition, a jitter of ± 35 ms around the nominal onset points was applied separately to each tone (i.e., the tones were shifted forward or backward in time).

Onset time, frequency, and location were co-varied in nine discrete steps covering the ranges specified above with linear spacing. The resulting smallest step of the jitter on each feature dimension was at or somewhat below the just-noticeable difference to allow for nine different steps while keeping the overall ranges of the feature values at manageable levels. The middle (fifth) step thus corresponded to the nominal value, whereas steps 1 to 4 produced tones that were slightly later, lower in frequency, and left-lateralized, and steps 6 to 9 produced tones that were slightly earlier, higher in frequency, and right-lateralized. For each individual tone, one of the nine steps was chosen, and this choice determined all three feature values.

In the conditions with jitter (conditions 1 to 4), each of the steps (1-9) occurred equally often to ensure that the amount of long-term (overall) feature variation (i.e., over the course of the whole sequence) was identical across conditions. Note that the amount of short-term feature variation (i.e., from one tone to the next) nevertheless differed between conditions; this will be discussed below. The assignment of the steps to the individual tones was either chosen randomly or followed pre-defined regular patterns. In condition 1 (*no perceptual organization predictable*), the order of the steps was entirely random.

In condition 2 (*segregated organization predictable*), the steps changed in an ordered manner across subsequent tones independently within each set (‘A’ and ‘B’). Separately for the ‘A’ and ‘B’ tones, the steps were slowly progressing from 1 to 9 and back (i.e.,

1234567898765432123...). This corresponds to separately introducing sequential predictability between successive 'A' tones ('A1–A2' and 'A2–A1', where 'A1' and 'A2' refer to the first and second 'A' within the same 'ABA-' triplet) and between successive 'B' tones ('B–B'). Because the order was progressing twice as fast for the 'A' as for the 'B' tones (due to the higher stimulation rate), the regularity in the overall ('ABA-') cycle spanned 9.6 seconds, thus it was unlikely to become a strong grouping cue.

In condition 3 (*integrated organization predictable*), the three tones within an 'ABA' triplet were assigned the same step value, whilst the step for the next triplet was chosen randomly. This led to a pattern of the type '222-444-999-111-777...' for the overall ('ABA-') cycle. Hence, the feature value of the 'A1' tone was predictive of the feature value of the subsequent 'B' tone, which in turn was predictive of the subsequent 'A2' tone. Unavoidably, the 'A1' tone was also predictive of the 'A2' tone within the same triplet. Importantly, however, 'A2' was never predictive of the next 'A1' tone. In terms of the integrated percept, this means that two out of the three links included in this organization would be strengthened ('A1–B' and 'B–A2', but not 'A2–A1'). In terms of the segregated percept, only every other link would be strengthened for the 'A' tone set ('A1–A2', but not 'A2–A1'), and no sequential link would be strengthened for the 'B' tone set.

In condition 4 (*integrated and segregated organizations predictable*), the three tones within an 'ABA' triplet were assigned the same step value (as in condition 3), and the step for the next triplet was chosen based on a regular progression from 1 to 9 and back (as in condition 2). This led to a pattern of the type '111-222-333-444-555...' for the overall ('ABA-') cycle. In this condition, sequential predictability was present for all possible pairs of tones.

In condition 5 (*no-variation control*), the nominal value of each feature (i.e., step 5) was chosen for each tone. Audio example files for each of the five conditions are available online at http://www.uni-oldenburg.de/fileadmin/user_upload/psycho/ag/psychophysio/audio/.

Procedure. Participants were instructed to listen to the tone sequences and to continuously indicate their percepts by depressing one of three specified buttons on the response pad (lower left,

lower right, or upper right). The fourth (upper left) button served to initiate the stimulus block. Participants were not made aware of the presence of the predictable patterns in the different conditions; they were simply asked to fulfill the same listening task for each stimulus block. For marking their percepts, participants were instructed to choose between four response alternatives: *Integrated* (depress one button) when they perceived the low and high tones as one coherent stream, *Segregated* (depress another button) when they perceived a low and a high stream in parallel, *Both* (depress a third button) when they perceived a stream consisting of low and high tones and an additional separate stream consisting of only low or only high tones, and *Neither* (release all buttons) when their current percept did not fall into any of these categories. The assignment of *Integrated*, *Segregated* and *Both* responses to the three buttons was counterbalanced across participants. Participants were encouraged to employ a neutral listening set, refraining from attempting to perceive the sequences according to one or another organization. The experimenter made sure that participants understood the types of percepts they were required to report using both auditory and visual illustrations.

The order of the five conditions (blocks) was separately randomized for each participant. A break of at least 30 seconds separated successive stimulus blocks, with additional time given to the participant as needed. Before the first experimental block, 1-min practice blocks were given as long as needed to clarify the instructions and the button-response assignment. Stimuli in the training block were chosen according to condition 1 (*no perceptual organization predictable*) in order to familiarize participants with the feature variation while not pre-exposing them to the regular patterns.

Data recording and analysis. The state of the three response buttons was continuously recorded with a sampling rate of 250 Hz. Before analyzing the button presses, all cases in which the duration between successive responses was shorter than 300 ms were discarded because these were assumed to result from inaccuracies in synchronizing the button presses and releases [34]. After this correction (which on average led to the exclusion of 0.62% of the responses), perceptual *phases*

were extracted from the participants' button presses. A perceptual phase is thus defined as the perception of the same sound organization for more than 300 ms.

Ten perceptual measures were derived separately for each participant and condition. The *proportion of Integrated* denotes the percentage of time in which an 'Integrated' percept was reported. The *mean duration of Integrated* indicates the average duration of 'Integrated' perceptual phases. Proportions and mean durations for *Segregated*, *Both* and *Neither* percepts were defined in an analogous manner. In addition, the *latency of the first Integrated percept* (the time from the start of the tone sequence to the onset of the first integrated perceptual phase) and the *latency of the first Segregated percept* (defined in an analogous manner) were determined. These latencies reflect the time it takes for the integrated or segregated organization to become the dominant percept for the first time. This period may or may not contain other perceptual phases. Because the 'Both' response very rarely appears as the first reported percept in a stimulus block and the 'Neither' response is indistinguishable from the delay to the first reported response, the latencies of the first integrated and segregated percept fairly well capture the dynamics of perception at the beginning of the stimulus block. Note that because each participant in each condition only records a single first percept, it is not possible to statistically analyze first percepts separately for 'Integrated' and 'Segregated'.

All duration analyses were carried out with log-transformed duration values to accommodate skewed distributions of the duration measures. The log scale was converted back to seconds for reporting mean durations and for display purposes.

Because the proportion of 'Neither' percepts was very low (0.69% across conditions; range 0.15% to 1.27%) and because neither the proportion nor the mean duration of 'Neither' percepts were affected by condition (all p values $> .05$), these two variables were excluded from further analysis. Condition effects were assessed for each of the remaining eight dependent measures (proportions of 'Integrated', 'Segregated', and 'Both' percepts; mean durations of 'Integrated', 'Segregated', and 'Both' percepts; latencies of the first 'Integrated' and first 'Segregated' percepts).

First, a control analysis was performed to examine the effects of introducing feature variation on the perceptual reports. To this end, conditions 4 and 5 were compared by means of two-tailed, paired-sample Student's t tests. Note that conditions 4 and 5 are identical with respect to predictability (they are regular both in the integrated and in the segregated organization) and differ only in the presence versus absence of feature variation.

For comparison with previous studies that did not distinguish between the amount of feature variation and the predictability of the features, a second control analysis was performed to assess the combined effects of introducing feature variation and unpredictability. To this end, conditions 1 (random feature variation, thus unpredictable) and 5 (no variation, thus predictable) were compared by means of two-tailed, paired-sample Student's t tests.

The main experimental analysis held the amount of long-term feature variation constant (by using only those conditions with feature variation, 1 to 4) and assessed the effects of introducing predictability to the tone transitions belonging to the integrated percept, to the segregated percept, or to both the integrated and segregated percepts (compared to the lack of predictability in the baseline condition). This was done in a repeated-measures analysis of variance (ANOVA) with the factors *predictability of the integrated organization* (2 levels: predictable vs. unpredictable) and *predictability of the segregated organization* (2 levels: predictable vs. unpredictable). In this analysis, conditions 1/4 correspond to the cases where both types of predictability are absent/present, condition 2 corresponds to integrated unpredictable/segregated predictable, and condition 3 corresponds to integrated predictable/segregated unpredictable.

Results

Figure 2 shows the group-average proportions of the 'Integrated', 'Segregated', 'Both' and 'Neither' percepts for the five different conditions. Figure 3 shows the group-average mean phase durations of the four percepts in each condition. Figure 4 shows the group-average latencies of the first 'Integrated' and 'Segregated' percepts in each condition. All three figures suggest that

perceptual reports were numerically affected by the experimental manipulations, consistent with the hypotheses.

--- insert Figures 2, 3, 4 at about here ---

Perceptual reports in condition 4 (with feature variation, but predictable in the integrated and segregated organizations) were slightly shifted towards stream integration when compared with condition 5 (without feature variation, thus also predictable in the two organizations). The control analysis comparing these conditions (cf. Table 1) showed that none of these differences reached statistical significance (all p values $> .05$), although some tendencies were observed (some p values < 0.10). These results suggest that the introduction of feature variation without changing the predictability of the sequence had relatively small effects.

--- insert Table 1 at about here ---

Perceptual reports in condition 1 (with unpredictable feature variation) were slightly shifted away from stream integration when compared with condition 5 (without feature variation). The control analysis comparing these conditions (cf. Table 2) again showed that none of the differences reached statistical significance (all p values $> .05$), although again some tendencies were observed (some p values < 0.10). These results suggest that the introduction of random feature variation (i.e., jointly changing the amount of feature variation and the predictability of the sequence) also had relatively small effects. Notably, introducing predictable (condition 4) or random (condition 1) feature variation tended to cause effects in opposite directions relative to the no-variation control condition (5).

--- insert Table 2 at about here ---

In the main experimental analysis, the amount of long-term feature variation was held constant, and the effects of introducing predictability were examined. Table 3 summarizes the results of the two-factorial ANOVAs assessing the effects of introducing predictability to the tone transitions belonging to the integrated percept (*integrated organization predictable*) and to the segregated percept (*segregated organization predictable*) as well as their interaction.

--- insert Table 3 at about here ---

Introducing predictability into the integrated organization had pronounced effects on almost all dependent measures: It increased the proportion of integrated reports and decreased the proportion of segregated reports; it prolonged the duration of integrated perceptual phases and shortened the duration of segregated phases; it led to an earlier discovery of the integrated percept and to a later discovery of the segregated percept.

Introducing predictability into the segregated organization also affected the perceptual reports, although the effects were less pronounced than those of predictability within the integrated organization. Predictability of the segregated organization significantly increased the proportion of segregated reports, but decreased the proportion of integrated reports only by tendency ($p = .10$). The main effect on the proportion of segregated reports was qualified by an interaction of the effects of integrated and segregated predictability: When the integrated organization was predictable, there was no significant effect of additionally making the segregated organization predictable (comparison of conditions 3 and 4, $t(19) = 1.162$, $p = .260$). In contrast, when the integrated organization was unpredictable, making the segregated organization predictable significantly increased the proportion of segregated reports (comparison of conditions 1 and 2, $t(19) = 3.717$, $p < .01$).

Introducing predictability into the segregated organization did not affect the mean duration of integrated or segregated perceptual phases, but it led to an earlier discovery of the segregated percept. Furthermore, it reduced the proportion and the mean duration of 'Both' perceptual phases. These main effects on the 'Both' percept were qualified by an interaction of the effects of integrated and segregated predictability. This interaction was caused by the fact that perceptual reports of 'Both' were enhanced only in one specific condition: When neither the integrated nor the segregated organization were predictable (condition 1), there was a higher proportion of 'Both' percepts (12%) than in any other condition as determined by pairwise follow-up t tests (condition 2: 3%, $p < .05$; condition 3: 2%, $p = .069$; condition 4: 1%, $p < .05$). There was a tendency towards a

similar pattern of results for the mean duration of ‘Both’ percepts (condition 1 vs. 2, $p = .056$; condition 1 vs. 3: $p = .096$; condition 1 vs. 4: $p < .01$).

In summary, introducing predictability into the integrated organization clearly promoted the perception of stream integration. Introducing predictability into the segregated organization supported the perception of stream segregation, though stronger effects were observed in the absence than in the presence of additional integrated predictability. When neither the integrated nor the segregated organizations were predictable, perceptual reports of ‘Both’ increased.

Discussion

The present study was designed to investigate the effects of predictability on auditory stream integration and segregation in an ‘ABA–’ tone paradigm. This was done by independent manipulations of predictability in the overall ‘ABA–’ sequence and separately in the ‘A’ and ‘B’ subsequences. In line with the hypotheses, predictability in the overall ‘ABA–’ sequence supported stream integration, whereas separate predictability in the ‘A’ and ‘B’ subsequences supported stream segregation. These effects were so pronounced as to change perceptual reports from 62% Integrated / 35% Segregated (condition 3, *integrated organization predictable*) to 26% Integrated / 73% Segregated (condition 2, *segregated organization predictable*). In other words, perception qualitatively flipped from predominantly integrated to predominantly segregated. This demonstrates a huge dynamic range within which predictability can affect perceptual organization, at least for a balanced combination of frequency separation and presentation rate. Note that feature similarity between the ‘A’ and ‘B’ tones, otherwise the strongest determinant of stream segregation versus integration [2,3,7], did not change at all between the different predictability conditions.

The independent manipulation of predictability was possible by systematically changing the sequential linkage between successive tones within each set (‘A–A–A’ and ‘B–B–B’) or across the sets (‘A–B’ and ‘B–A’). Tones that are predictive of each other can be assumed to be more strongly linked in perception, and thus should be more likely to form a coherent stream [16–18,33]. This

idea had been empirically demonstrated by introducing predictability separately into the ‘A’ and ‘B’ tone sets and showing that this increases stream segregation relative to an unpredictable control condition [5,30]. The present study replicates these earlier findings (cf. the contrast of conditions 1 and 2). The converse case, in which predictability would be introduced into the overall ‘ABA–’ sequence without making the ‘A’ and ‘B’ subsequences predictable, had not yet been empirically tested. The novel contribution of the present study lies in the demonstration that making the overall sequence predictable enhances stream integration relative to an unpredictable control condition (cf. the contrast of conditions 1 and 3). This result further supports the view that predictive relations between tones increase their sequential linkage and thus their perceptual coherence in an auditory stream ([16–18,33]; see also early work by [6,19,21,22]).

Previous studies [5,30] have also attempted to delineate how the effect of predictability on the distribution of the perceptual reports was brought about. They did this by analyzing the average phase durations for the different perceptual states as a measure of perceptual stability. Results suggested that introducing predictability separately into the ‘A’ and ‘B’ subsequences increases the proportion of ‘Segregated’ perceptual reports only by prolonging segregated percepts, but not by shortening integrated percepts (cf. [35]). This was taken to imply that predictability is only evaluated for the currently dominant perceptual organization [5,30]. Hence, predictability was considered a higher-order cue that comes into play only after a perceptual decision has been made, and that is acting only to stabilize the resulting percept [35].

Results of the present study are not as clear-cut as the previous ones in this respect. The effect of predictability within the integrated organization on the proportion of integrated and segregated perceptual reports was caused by prolonging (i.e. stabilizing) integrated percepts, but also by cutting short segregated percepts (i.e., causing switches back to integration). This pattern of results is more consistent with that of a primary cue that affects the initial perceptual decision between stream integration and segregation [5,30,35,36]. The effect of predictability within the segregated organization on the proportion of segregated and integrated perceptual reports could not

be attributed to a change in duration of the segregated or integrated percept (neither of these effects were significant). Instead, it was mainly caused by an earlier discovery of the segregated percept in the presence of segregated predictability. Thus the segregated percept entered the dynamic competition earlier, and thereby could be experienced for a larger proportion of the time than in the absence of segregated predictability. This pattern is also suggestive of segregated predictability acting as a primary rather than a higher-order cue [5,30,35,36]. This finding qualitatively differs from those of previous regularity manipulation experiments [5,30]. The reason for this discrepancy is unclear: It might stem from differences in the employed regularities or in the features that were carrying the regularities (multiple co-varying features in the present study). Future studies should explicitly contrast the effects of different types of feature regularities and investigate whether this affects the stage at which regularities exert their influence (i.e., primary vs. higher-order effect). This is particularly relevant as temporal predictability may be qualitatively different from predictability in other feature dimensions such as frequency and location [37].

Another possibility is that the complexity of the predictability manipulation determines whether primary or higher-order effects are obtained. Although the present way of manipulating predictability was not based on simple repetition, it was still less complex than the regular patterns used in some previous studies associated with higher-order effects of the predictability manipulation [5,30]. Using smooth progression between the feature values causes smaller short-term variation of the features for the predictable compared to the unpredictable conditions.. It is possible that “feature-shift” detectors picked up such gradual changes [38], providing information for auditory scene analysis processes without the need for extracting predictability. Future studies should thus implement predictability manipulations in which the characteristics of tone-to-tone transitions (i.e., short-term feature variation) are identical between the predictable and unpredictable conditions. This would also be important for further distinguishing between genuine effects of predictability and the Gestalt principle of continuity [1].

Notwithstanding these limitations, the present study adds to the growing body of evidence on predictability as a cue within auditory scene analysis [4,5,28–32], but over and above that, it offers a generalization of previous approaches. Each of the manipulations employed in previous studies can be mapped onto one of the present manipulations.

The present comparison of conditions 1 (*no perceptual organization predictable*) and 5 (*no-variation control*) corresponds to the approach of making the whole sequence unpredictable by jittering some feature values of the ‘A’ and ‘B’ tones and comparing the perception of such sequences to those with a constant (thereby predictable) arrangement ([8] for frequency; [23] for timing; [24] for timing in a biasing introduction sequence). As noted above, this entails a parallel manipulation of feature variation and predictability; moreover, predictability of the integrated and the segregated perceptual organization are affected in parallel. Because these co-varying factors partly act in opposite directions, it is not surprising to see that the effects of this manipulation are relatively low both in the present study (cf. Table 2) and in previous studies [8,23,24]. These results, although informative e.g. for computational model of auditory scene analysis, are not conclusive concerning the role of predictability as a cue for stream segregation or integration.

The present comparison of conditions 1 (*no perceptual organization predictable*) and 4 (*integrated and segregated organizations predictable*) corresponds to the approach of comparing a randomly jittered version of the ‘ABA–’ / ‘ABAB’ sequence with an equal overall amount of jitter but arranged in predictable patterns ([39] for the succession of the ‘A’ and ‘B’ tones; [23] for slight variations in the ‘A’ and ‘B’ frequency values). Although the amount of long-term variation is controlled for in this comparison, the approach still leads to a parallel manipulation of predictability for the integrated and segregated perceptual organizations. Any resulting effects on stream integration and segregation would again act in opposite directions. This explains why no effects on perception were caused by this manipulation in previous studies [23,39]. In the present study, condition 4 was perceived as somewhat more integrated than condition 1; this is consistent with a generally larger effect of integrated than segregated predictability in the present dataset.

Nevertheless, due to the confounding parallel influence on integrated and segregated predictability, these results are again not informative as to the role of predictability within sequential auditory scene analysis.

The present comparison of conditions 1 (*no perceptual organization predictable*) and 2 (*segregated organization predictable*) corresponds to the approach of some more recent studies comparing an unpredictable arrangement (random jitter) with a condition in which only one of the streams was arranged in a predictable pattern ([4,28,29]; some conditions in [5]) or in which the ‘A’ and ‘B’ streams received independent predictable patterns [5,30]. As noted above, all these studies have reached the same conclusion as the present one: A selective enhancement of the predictability of the segregated organization leads to an increase in the perception of stream segregation.

The present results suggest that this finding reflects only part of the picture: The opposite manipulation, selectively enhancing the predictability of the integrated organization, leads also to a strong opposite effect, namely a notable increase in the perception of stream integration. This is shown by the present condition 3 (*integrated organization predictable*) that leads to a higher proportion of integrated percepts than any other condition. Together, the effects of selectively enhancing integrated versus segregated predictability were three to four times larger than the ones previously observed when manipulating segregated predictability alone [5,30]. These pronounced predictability effects resulted from *directional* predictability – i.e., predictive relations that favor either integration or segregation, but not both. In information-theoretic terms, condition 4 is more predictable than any other condition; yet it does not cause the strongest effects on perception. Similarly, condition 1 is less predictable than any other condition; yet again, its effects fall between those of the other conditions. The “extreme points” on a continuum of predictability manipulation are thus represented by the conditions with only integrated or only segregated predictability (cf. the arrangement of conditions on Figures 2, 3 and 4). By conceptualizing other manipulations (integrated and segregated predictability / neither integrated nor segregated predictability) as intermediate points on this continuum, the present framework can embrace previous investigations

and explain why their effects have proven less robust. Future studies attempting to further specify the role of predictability within auditory scene analysis should therefore focus on *directional* predictability manipulations in order to avoid the interpretational ambiguities arising from a parallel manipulation of integrated and segregated predictability.

Although not apparent at first sight, the results obtained in a recent study [31] also fit the present framework. In this study, a condition without jitter (i.e. predictable via constancy, corresponding to present condition 5) was compared against a condition with temporal jitter in the ‘B’ stream. Such temporal jitter removes predictability from the integrated organization, while the segregated organization becomes unpredictable for one of the streams but remains predictable for the other stream. The jittered condition of [31] thus reflects a case of partial segregated predictability that can be considered to lie in between conditions 1 (*no organization predictable*) and 2 (*segregated organization predictable*) in the present conceptualization. Consistent with the present results, perceptual reports were more segregated in this jittered condition than in the constant (predictable) control condition. The authors interpret their finding from the opposite point of view, emphasizing that perceptual reports become more integrated in the predictable condition. Both views appear equally valid; they essentially depend on the definition of a neutral reference point (against which to compare the effects of a given predictability manipulation).

One further aspect of the present dataset also fits the notion that predictability acts as a directional cue. In condition 1 (*no perceptual organization predictable*), neither the integrated nor the segregated perceptual organizations would have been particularly favored for being perceptually coherent. This condition was associated with a specific increase in the occurrence of a third perceptual alternative, the ‘Both’ percept. ‘Both’ denotes a perceptual category including patterns made from one stream containing low and high tones and an additional separate stream containing only low or only high tones (e.g., perceiving one stream of ‘AB--AB--AB--’ and a parallel stream of ‘--A---A---A-’, or perceiving one stream of ‘-BA--BA--BA-’ and a parallel stream of ‘A---A---A---’; cf. [40]). Not many previous studies have taken into account the possibility that there might

be other perceptual alternatives beyond the classical integrated and segregated patterns described by [7]; hence listeners were usually not given the opportunity to denote the ‘Both’ percept by a corresponding response button. Consequently, not much is known about the conditions under which the ‘Both’ percept occurs, other than that it appears more rarely than the integrated and segregated alternatives, and usually enters perceptual competition later than those [9,40]. The present finding that the ‘Both’ percept becomes more frequent when neither the integrated nor the segregated perceptual organizations are predictable can be interpreted in the framework of perceptual coherence [16–18,33]: When the two predominant alternatives (integrated and segregated) do not provide “satisfactory” explanations of the sound input due to the absence of predictability, perception may start exploring other grouping alternatives.

Because of the increase in perceptual reports of ‘Both’, condition 1 (*no perceptual organization predictable*) is not a fully neutral condition that lies in between the conditions with only integrated or only segregated predictability. In fact, it is difficult to infer from the present dataset which condition is most suited as a neutral comparison point for future studies employing directional predictability manipulations. Although the differences between conditions 1 (*no organization predictable*), 4 (*integrated and segregated organizations predictable with jitter*) and 5 (*integrated and segregated organizations predictable without jitter*) were not particularly strong (cf. Tables 1 and 2), they were also not negligible. From a theoretical point of view, the classical version of the ‘ABA-‘ paradigm with constant feature values is least preferable because any reduction of predictability relative to this condition necessarily confounds effects of feature variation and predictability. Notably, departing from this constant-feature condition by introducing predictable feature variation (condition 4) or random feature variation (condition 1) caused small effects in opposite directions in the present study. Future studies should bear this in mind and ideally implement both control conditions so as to obtain a comprehensive picture.

Regardless of whether the present conditions 1, 4, or 5 are considered as the most appropriate baseline, condition 3 clearly shows an effect of integration beyond baseline, and

condition 2 an effect of segregation beyond baseline. Besides the appeal of being able to bias perceptual reports in either direction, this finding has important theoretical implications for conceptualizing the process of auditory scene analysis. A long-standing debate concerns the issue of whether integration represents a “default state” of the auditory perceptual system when exposed to an ambiguous sequence (cf. discussion in [41–44]). The “default state” interpretation suggests that the system collects evidence only in favor of stream segregation, and whenever the collected evidence is not sufficient, sounds are perceived as coming from one sound source (integrated). A contrary view posits that stream integration is also an active process that requires the collection of perceptual evidence (e.g., [18,45]). The present results show that enhancing the predictability of the integrated organization biases perception towards integration independently of the presence or absence of predictability in the segregated organization. This suggests that integration is not a “default mode” of organizing sequential input: The system is also considering cues that favor stream integration. This suggests that all organizations are “equal” in that they must be supported by cues present in the auditory input. This conclusion is consistent with other recent work challenging the view of “integration by default” [9,46].

In conclusion, the present study demonstrates that sequential predictability can flexibly support either the integrated or the segregated perception of an auditory sequence when it is set up as a directional cue, selectively affecting predictability of only the integrated or only the segregated perceptual organization. Furthermore, results suggest a continuum between these extreme points along which previous studies can be grouped to understand the presence, direction and strength of the effect of a given predictability manipulation on auditory perception.

References

1. Bregman AS. Auditory scene analysis. The perceptual organization of sound. Cambridge, MA: MIT Press; 1990.
2. Moore BCJ, Gockel HE. Factors influencing sequential stream segregation. *Acta Acust United Acust.* 2002;88:320–333.

3. Moore BCJ, Gockel HE. Properties of auditory stream formation. *Philos Trans R Soc Lond B Biol Sci.* 2012;367(1591):919–931.
4. Andreou L-V, Kashino M, Chait M. The role of temporal regularity in auditory segregation. *Hear Res.* 2011;280(1-2):228–235.
5. Bendixen A, Denham SL, Gyimesi K, Winkler I. Regular patterns stabilize auditory streams. *J Acoust Soc Am.* 2010;128(6):3658–3666.
6. Jones MR. Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychol Rev.* 1976;83(5):323–355.
7. Van Noorden LPAS. Temporal Coherence in the Perception of Tone Sequences. Doctoral dissertation. Eindhoven, The Netherlands: Technical University Eindhoven; 1975.
8. Denham SL, Gyimesi K, Stefanics G, Winkler I. Stability of Perceptual Organisation in Auditory Streaming. In: Lopez-Poveda EA, Palmer AR, Meddis R, Editors. *Neurophysiol Bases Audit Percept.* New York, NY: Springer; 2010. pp. 477–487.
9. Denham SL, Gyimesi K, Stefanics G, Winkler I. Perceptual bistability in auditory streaming: How much do stimulus features matter? *Learn Percept.* 2013;5(S2):73–100.
10. Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ. Neuromagnetic Correlates of Streaming in Human Auditory Cortex. *J Neurosci.* 2005;25(22):5382–5388.
11. Hill KT, Bishop CW, Yadav D, Miller LM. Pattern of BOLD signal in auditory cortex relates acoustic response to perceptual streaming. *BMC Neurosci.* 2011;12(1):85.
12. Kondo HM, Kashino M. Involvement of the Thalamocortical Loop in the Spontaneous Switching of Percepts in Auditory Streaming. *J Neurosci.* 2009;29(40):12695–12701.
13. Pressnitzer D, Hupé J-M. Temporal Dynamics of Auditory and Visual Bistability Reveal Common Principles of Perceptual Organization. *Curr Biol.* 2006;16(13):1351–1357.
14. Schwartz J-L, Grimault N, Hupé J-M, Moore BCJ, Pressnitzer D. Multistability in perception: binding sensory modalities, an overview. *Philos Trans R Soc Lond B Biol Sci.* 2012;367(1591):896–905.
15. Szalárdy O, Böhm TM, Bendixen A, Winkler I. Event-related potential correlates of sound organization: Early sensory and late cognitive effects. *Biol Psychol.* 2013;93(1):97–104.
16. Denham SL, Winkler I. The role of predictive models in the formation of auditory streams. *J Physiol-Paris.* 2006;100(1-3):154–170.
17. Winkler I, Denham SL, Nelken I. Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn Sci.* 2009;13(12):532–540.
18. Winkler I, Denham SL, Mill RW, Böhm TM, Bendixen A. Multistability in auditory stream segregation: a predictive coding view. *Philos Trans R Soc Lond B Biol Sci.* 2012;367(1591):1001–1012.
19. Jones MR, Boltz M. Dynamic attending and responses to time. *Psychol Rev.* 1989;96(3):459–491.

20. Drake C, Jones MR, Baruch C. The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition*. 2000;77(3):251–288.
21. Jones MR, Kidd G, Wetzel R. Evidence for rhythmic attention. *J Exp Psychol Hum Percept Perform*. 1981;7:1059–1073.
22. Jones MR, Boltz M, Kidd G. Controlled attending as a function of melodic and temporal context. *Percept Psychophys*. 1982;32(3):211–218.
23. French-St. George M, Bregman AS. Role of predictability of sequence in auditory stream segregation. *Percept Psychophys*. 1989;46(4):384–386.
24. Rogers WL, Bregman AS. An experimental evaluation of three theories of auditory stream segregation. *Percept Psychophys*. 1993;53(2):179–189.
25. Baldeweg T. Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends Cogn Sci*. 2006;10(3):93–94.
26. Friston K. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci*. 2005;360(1456):815–836.
27. Winkler I. Interpreting the Mismatch Negativity. *J Psychophysiol*. 2007;21(3):147–163.
28. Rimmele J, Schröger E, Bendixen A. Age-related changes in the use of regular patterns for auditory scene analysis. *Hear Res*. 2012;289(1-2):98–107.
29. Devergie A, Grimault N, Tillmann B, Berthommier F. Effect of rhythmic attention on the segregation of interleaved melodies. *J Acoust Soc Am*. 2010;128(1):EL1–EL7.
30. Bendixen A, Böhm TM, Szalárdy O, Mill RW, Denham SL, Winkler I. Different roles of similarity and predictability in auditory stream segregation. *Learn Percept*. 2013;5(S2):37–54.
31. Rajendran VG, Harper NS, Willmore BD, Hartmann WM, Schnupp JWH. Temporal predictability as a grouping cue in the perception of auditory streams. *J Acoust Soc Am*. 2013;134(1):EL98–EL104.
32. Bendixen A, Schröger E, Ritter W, Winkler I. Regularity Extraction from Non-Adjacent Sounds. *Front Psychol*. 2012;3:143.
33. Mill RW, Böhm TM, Bendixen A, Winkler I, Denham SL. Modelling the Emergence and Dynamics of Perceptual Organisation in Auditory Streaming. *PLoS Comput Biol*. 2013;9(3):e1002925.
34. Moreno-Bote R, Shpiro A, Rinzel J, Rubin N. Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *J Vis*. 2010;10(11):1, 1–18.
35. Bendixen A, Denham SL, Winkler I. Sound predictability as a higher-order cue in auditory scene analysis. *AIA-DAGA 2013 Int Conf Acoust*. Berlin: DEGA; pp. 705–708.
36. Szalárdy O, Bendixen A, Tóth D, Denham SL, Winkler I. Modulation-frequency acts as a primary cue for auditory stream segregation. *Learn Percept*. 2013;5(S2):149–161.
37. Arnal LH, Giraud A-L. Cortical oscillations and sensory predictions. *Trends Cogn Sci*. 2012;16(7):390–398.

38. Demany L, Ramos C. On the binding of successive sounds: Perceiving shifts in nonperceived pitches. *J Acoust Soc Am*. 2005;117(2):833–841.
39. Carl D, Gutschalk A. Role of pattern, regularity, and silent intervals in auditory stream segregation based on inter-aural time differences. *Exp Brain Res*. 2013;224(4):557–570.
40. Denham SL, Böhm TM, Bendixen A, Szalárdy O, Kocsis Z, Mill R, Winkler I. Stable individual characteristics in the perception of multiple embedded patterns in multistable auditory stimuli. *Front Neurosci*. 2014;8:25.
41. Bregman AS. Auditory streaming: Competition among alternative organizations. *Percept Psychophys*. 1978;23(5):391–398.
42. Brochard R, Drake C, Botte M-C, McAdams S. Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *J Exp Psychol Hum Percept Perform*. 1999;25(6):1742–1759.
43. Cusack R, Deeks J, Aikman G, Carlyon RP. Effects of Location, Frequency Region, and Time Course of Selective Attention on Auditory Scene Analysis. *J Exp Psychol Hum Percept Perform*. 2004;30(4):643–656.
44. Thompson SK, Carlyon RP, Cusack R. An objective measurement of the build-up of auditory streaming and of its modulation by attention. *J Exp Psychol Hum Percept Perform*. 2011;37(4):1253–1262.
45. Sussman ES. Integration and segregation in auditory scene analysis. *J Acoust Soc Am*. 2005;117(3):1285–1298.
46. Deike S, Heil P, Böckmann-Barthel M, Brechmann A. The Build-up of Auditory Stream Segregation: A Different Perspective. *Front Psychol*. 2012;3:461.

Tables

Table 1: Results of the first control analysis assessing the effects of feature variation independently of predictability. Conditions 4 (with predictable feature variation) and 5 (without feature variation) were compared.

Dependent measure	Condition 4	Condition 5	Statistical comparison
Proportion of Integrated	56%	45%	$t(19) = 1.897, p = .073$
Proportion of Segregated	41%	51%	$t(19) = -1.725, p = .101$
Proportion of Both	3%	4%	$t(19) = -0.167, p = .869$
Mean duration of Integrated	8.46 s	7.88 s	$t(19) = 0.257, p = .800$
Mean duration of Segregated	6.56 s	10.91 s	$t(19) = -1.950, p = .066$
Mean duration of Both	0.49 s	0.66 s	$t(19) = -0.674, p = .508$
Latency of first Integrated percept	4.65 s	4.25 s	$t(19) = 0.590, p = .562$
Latency of first Segregated percept	27.81 s	16.85 s	$t(19) = 1.602, p = .126$

Table 2: Results of the second control analysis assessing the combined effects of introducing feature variation and unpredictability into the ‘ABA–’ sequence. Conditions 1 (with random feature variation) and 5 (without feature variation) were compared.

Dependent measure	Condition 1	Condition 5	Statistical comparison
Proportion of Integrated	34%	45%	$t(19) = -1.706, p = .104$
Proportion of Segregated	53%	51%	$t(19) = 0.297, p = .770$
Proportion of Both	12%	4%	$t(19) = 1.690, p = .108$
Mean duration of Integrated	5.03 s	7.88 s	$t(19) = -1.212, p = .241$
Mean duration of Segregated	10.01 s	10.91 s	$t(19) = -0.346, p = .733$
Mean duration of Both	1.43 s	0.66 s	$t(19) = 1.763, p = .094$
Latency of first Integrated percept	4.58 s	4.25 s	$t(19) = 0.217, p = .831$
Latency of first Segregated percept	16.05 s	16.85 s	$t(19) = -0.202, p = .842$

Table 3: Group-average mean values and statistical results of the main experimental analysis assessing the effect of predictability while holding the amount of long-term feature variation constant (based on conditions 1 to 4). Significant results are marked in bold font.

Dependent measure	Condition 3 (integrated organization predictable)	Condition 4 (integrated and segregated organizations predictable)	Condition 1 (no perceptual organization predictable)	Condition 2 (segregated organization predictable)
Proportion of Integrated	62%	56%	34%	26%
	Integrated organization predictable: $F(1,19) = 32.818$, $p < .001$ Segregated organization predictable: $F(1,19) = 2.984$, $p = .100$ Interaction: $F(1,19) = 0.339$, $p = .568$			
Proportion of Segregated	35%	41%	53%	73%
	Integrated organization predictable: $F(1,19) = 28.835$, $p < .001$ Segregated organization predictable: $F(1,19) = 9.306$, $p < .01$ Interaction: $F(1,19) = 5.584$, $p < .05$			
Proportion of Both	2%	3%	12%	1%
	Integrated organization predictable: $F(1,19) = 2.324$, $p = .144$ Segregated organization predictable: $F(1,19) = 6.596$, $p < .05$ Interaction: $F(1,19) = 3.905$, $p = .063$			
Mean duration of Integrated	8.51 s	8.46 s	5.03 s	3.62 s
	Integrated organization predictable: $F(1,19) = 8.872$, $p < .01$ Segregated organization predictable: $F(1,19) = 0.429$, $p = .520$ Interaction: $F(1,19) = 0.618$, $p = .442$			
Mean duration of Segregated	5.89 s	6.56 s	10.01 s	10.51 s
	Integrated organization predictable: $F(1,19) = 5.811$, $p < .05$ Segregated organization predictable: $F(1,19) = 0.177$, $p = .679$ Interaction: $F(1,19) = 0.029$, $p = .867$			
Mean duration of Both	0.57 s	0.49 s	1.43 s	0.34 s
	Integrated organization predictable: $F(1,19) = 0.900$, $p = .355$ Segregated organization predictable: $F(1,19) = 7.334$, $p < .05$ Interaction: $F(1,19) = 5.475$, $p < .05$			
Latency of first Integrated percept	3.70 s	4.65 s	4.58 s	7.58 s
	Integrated organization predictable: $F(1,19) = 4.432$, $p < .05$ Segregated organization predictable: $F(1,19) = 2.166$, $p = .157$ Interaction: $F(1,19) = 0.398$, $p = .536$			
Latency of first Segregated percept	39.11 s	27.81 s	16.05 s	8.66 s
	Integrated organization predictable: $F(1,19) = 36.519$, $p < .001$ Segregated organization predictable: $F(1,19) = 7.014$, $p < .05$ Interaction: $F(1,19) = 0.921$, $p = .349$			

Figure Captions

Figure 1: Stimulus sequence examples for the five experimental conditions of the present study. Time is represented on the X axis and frequency on the Y axis of all panels. Shading of the tones corresponds to the feature steps (1-9) used for the conjoint manipulation of timing, frequency, and location. Dotted lines indicate the presence of predictive relations between successive tones (i.e., the feature values of one tone are predictive of the feature values of the next tone in one or both perceptual organizations). For tones that are not connected with a dotted line, the succession of the feature values is random. See main text for details.

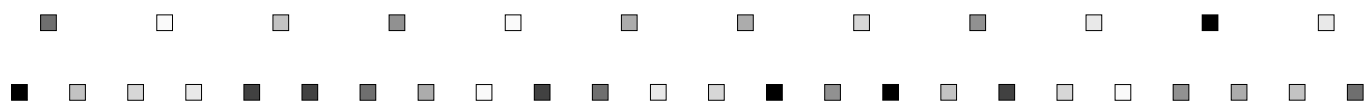
Figure 2: Group-average (N=20) proportions of ‘Integrated’ (Int), ‘Segregated’ (Seg), ‘Both’ (Bot) and ‘Neither’ (Nei) percepts as a function of experimental condition. Means and standard errors of mean are displayed.

Figure 3: Group-average (N=20) mean phase durations for ‘Integrated’ (Int), ‘Segregated’ (Seg), ‘Both’ (Bot) and ‘Neither’ (Nei) percepts as a function of experimental condition. Means and standard errors of mean are displayed. Duration data were log-transformed for the analyses. They are shown here after being converted back to seconds for display purposes.

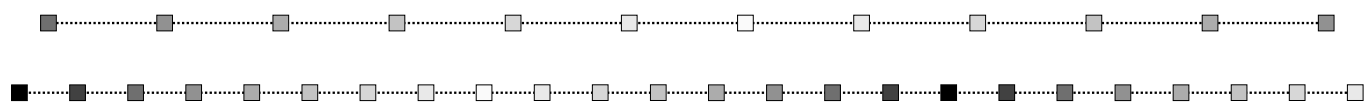
Figure 4: Group-average (N=20) latencies of the first ‘Integrated’ (Int) and ‘Segregated’ (Seg) percepts in the stimulus block as a function of experimental condition. Means and standard errors of mean are displayed. Latency data were log-transformed for the analyses. They are shown here after being converted back to seconds for display purposes.

Figure 1

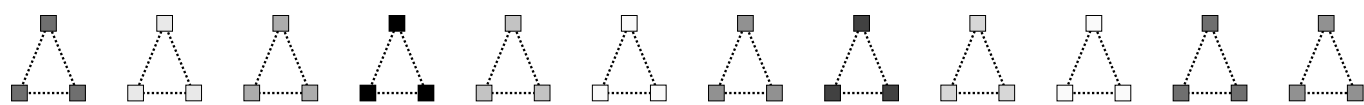
Condition 1:
No perceptual
organization
predictable



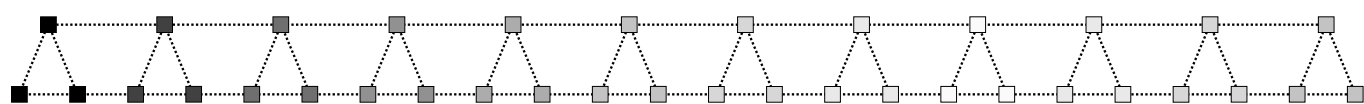
Condition 2:
Segregated
organization
predictable



Condition 3:
Integrated
organization
predictable



Condition 4:
Integrated and
segregated
organizations
predictable



Condition 5:
No-variation
control

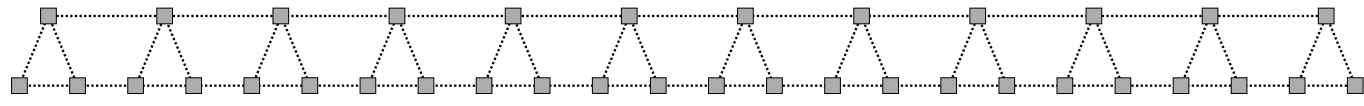


Figure 2

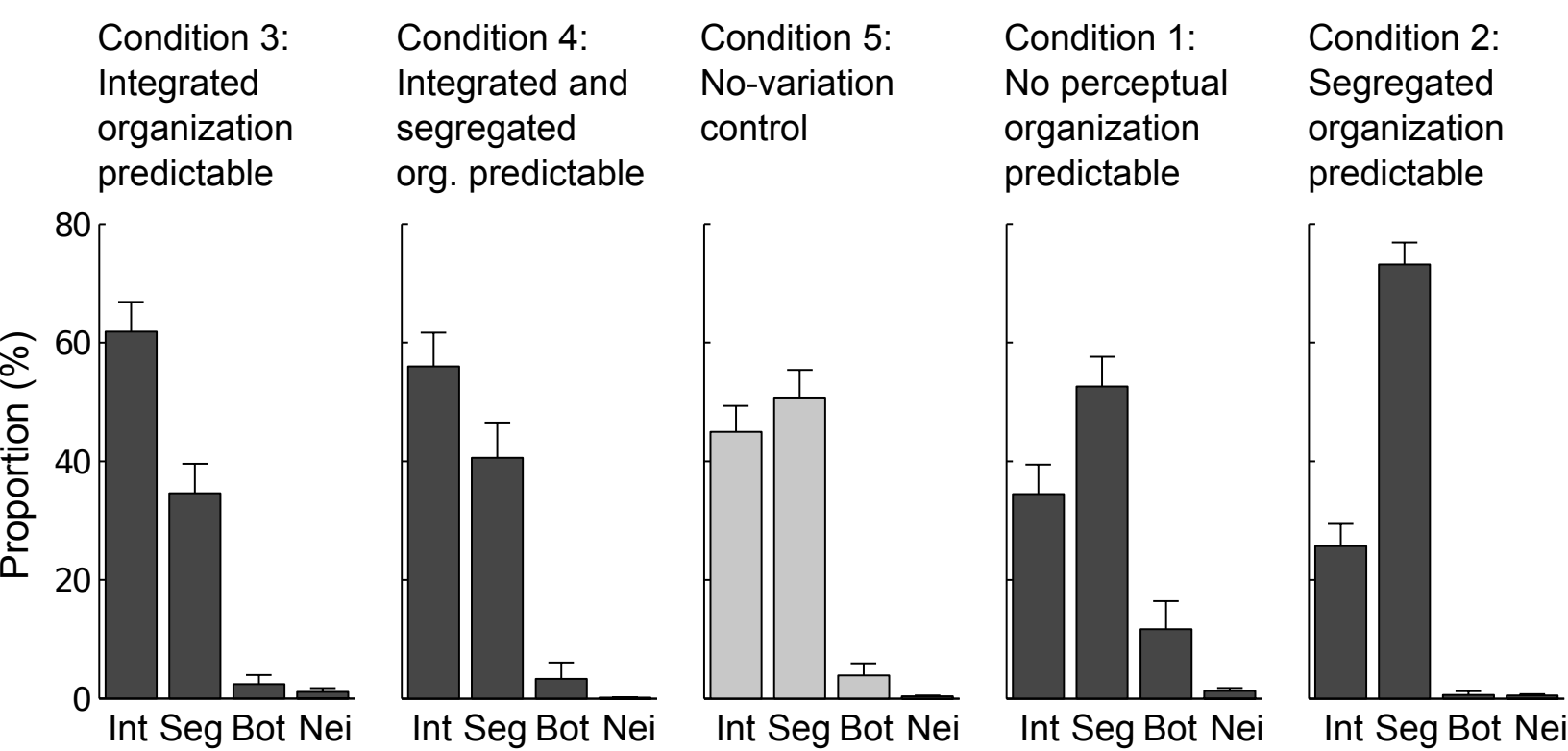


Figure 3

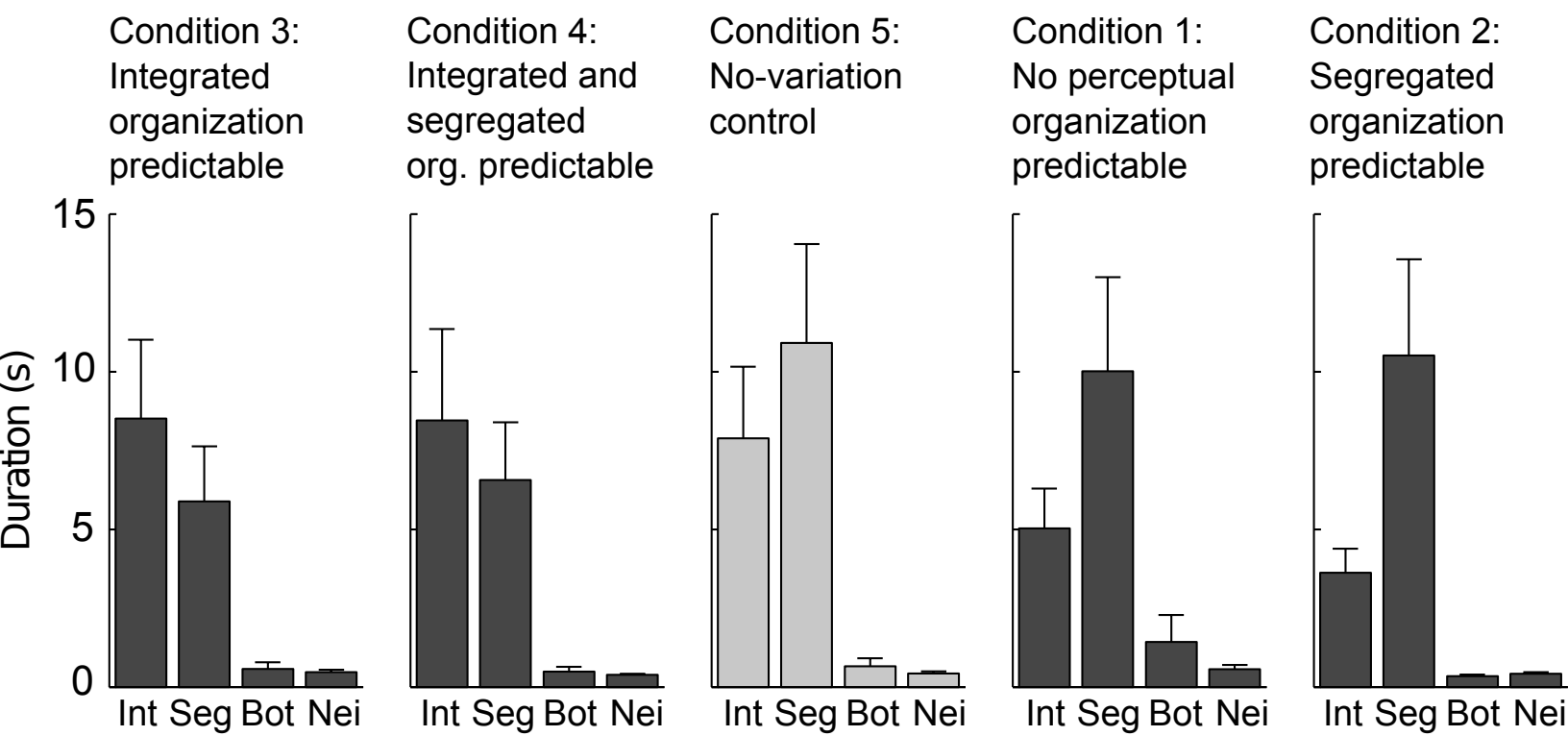


Figure 4

